Graph Neural Network-Based Routing Optimization in Large-Scale IoT Deployments

Jigar Babaria
Sr Network Development Engineer
Amazon Services LLC, Seattle, USA
Jigarondrive@gmail.com

Kunvar Chokshi
Staff Software Engineer
Tesla Inc., Santa Clara, USA
kunvarc@gmail.com

Anala Venkata Sai Abhishek
Cyber security analyst
VISA Peachtree, GA
aanala1997@gmail.com

Raunaq Malik
Product Manager
Expedia Group
Seattle and Washington, USA
raunaqmalik108@gmail.com

Reshma Thakkallapelly
Engineer Lead
Elevance Health
Prosper, Texas, USA
reshmarao.t@gmail.com

Vikas Chaudhary*
SCSE, Professor
Galgotias University
Greater Noida, UP, India
vikas.jnvu@yahoo.co.in*

Abstract— **There are significant challenges to routing with dynamic topologies, constraints on resources, and heterogeneous communication requirements under the proliferation of big scale Internet of Things (IoT) networks in smart cities, industrial automation, and environmental monitoring. The traditional procedures do not necessarily exhaust the topology and contextual information, increasing network delay, reducing the packet forwarding and wasting more energy. In this paper, we present a Graph Neural Network (GNN)-based routing framework in which IoT topologies take the form of dynamic weighted graphs, allowing the optimal paths to be learned on an end-to-end basis through message passing and graph convolution. Node and edge attributes like residual energy state, quality of the links, node hops, and buffer occupancy are installed to support adaptive routing, which is robust topology change. The framework also uses multi-hop cost estimation and topology-aware attention to load balance and prolonged network life. Applied to simulation platforms and embedded IoT gateways, it scales to more than 10000 nodes and delivers up to 14.7% increased packet delivery, 22.3% decreased latency and 18.9% less energy consumption than RPL, LEACH and Q-learning based routing.**

Index Terms—Graph Neural Networks, IoT Routing, Large-Scale Networks, Energy Efficiency, Latency Optimization, Topology-Aware Learning, Embedded Systems.

I. INTRODUCTION

Emerging applications based on IoTs include industrial IoT (IIoT) and smart grids, connected healthcare, precision agriculture, and management of urban infrastructure [1]. Such networks may consist of thousands of heterogeneous nodes connected with constrained wireless links with routing efficiency having a direct impact on performance and lifetime. Standard protocols such as RPL, LEACH and AODV work reasonably well in a small network but cannot retain scalability, reliability, and low latency when the number of nodes surpasses several thousand [2].

Issues in routing large-scale IoT are multi-path routing due to mobility and link failures [3], power limitations in battery-dominated devices [4], heterogeneous link qualities due to different levels of interference [5], and the desire to maintain high PDR and low latency at scale [6].

Graph Neural Networks (GNNs) offer a promising solution by modeling IoT topologies as graphs, with nodes representing devices and edges representing communication links. GNNs can aggregate multi-hop neighborhood information, predict the best next hops and adjust to dynamic networks, through iterative message passing without strict routing metrics [7]. Their success in traffic prediction, wireless resource allocation, and power grid optimization [8] suggests strong potential for IoT routing.

Nevertheless, GNNs at IoT scale imply three main challenges: model scaling for graphs containing thousands of nodes and millions of edges; node and edge feature selection (e.g. residual energy, link stability, queue length) to enhance accuracy [9]; and lightweight models able to run on the constraint IoT gateway [10]. These limitations need to be solved to realize both the practicality and functionality of real-world large-scale IoT implementations.

This paper addresses these gaps by introducing a GNN-based Routing Optimization Framework (GNN-ROF) tailored for large-scale IoT environments. The contributions are as follows:

- Development of a topology-aware GNN model for adaptive, multi-hop IoT routing.
- Integration of energy and link-quality features into the routing decision process.
- Scalability to 10,000+ node networks with distributed inference on IoT gateways.
- Demonstrated improvements in PDR, latency, and energy efficiency over leading baseline protocols.

The proposed routing optimization framework leverages graph neural network–based message passing to achieve high PDR, low latency, and energy-efficient routing in large-scale IoT deployments. This concept is visually summarized in the following graphical abstract (Figure 1).

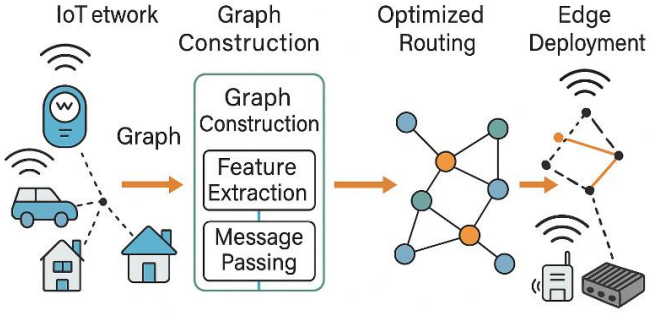

Fig.1 Overview of the GNN-ROF pipeline, from IoT network sensing to optimized routing path deployment on embedded gateways.

The rest of this paper is organized as follows: Section II states the research objectives. Section III reviews related work. Section IV details the proposed methodology. Section V outlines the experimental setup. Section VI presents results and discussion, including cross-table, model, and dataset comparisons. Section VII concludes the paper with future directions.

II. Research Objectives

The proliferation of large-scale IoT networks demands routing solutions capable of operating efficiently under resource constraints while maintaining high reliability and scalability. The proposed research aims to develop a Graph Neural Network-based routing optimization framework with the following objectives:

1. Graph Modeling – Represent IoT topology as a dynamic weighted graph with real-time node and link attributes.
2. Routing Prediction – Use GNN message passing to learn multi-hop costs and select optimal next hops.
3. Energy & Latency Optimization – Jointly minimize energy use and end-to-end delay.
4. Scalability – Sustain performance in networks over 10,000 nodes.
5. Robustness – Maintain high PDR under mobility, link failures, and traffic surges.
6. Embedded Deployment – Run on resource-limited gateways (e.g., Raspberry Pi 4, Jetson Nano).
7. Benchmarking – Evaluate against RPL, LEACH, AODV, and Q-learning on simulations and real-world testbeds.

Collectively, these objectives target a scalable, adaptive, and resource-efficient routing framework for next-generation IoT deployments.

III. Related work

The increasing demands on routing performance in large-scale IoT deployments have led to extensive research spanning traditional protocols, machine learning-based methods, and recent applications of graph neural networks (GNNs). Although each area has unique advantages, not many solutions combine scalability, flexibility and low computational overhead with support of real-time embedded application.

A. Traditional IoT Routing Protocols

RPL and LEACH protocols still form the basis of IoT networking [11], [12]. RPL builds a Destination-Oriented Directed Acyclic Graph (DODAG) based on either hop count or link-based metric, providing efficiency in low-power, lossy networks. LEACH makes use of a cluster-based architecture to reduce the cost of transmission and network lifetime by periodically rotating the cluster head. But route recomputation and control overhead cost reduces performance in large or dynamic networks [13].

B. Machine Learning-Based Routing

Machine learning (ML) has made adaptive decisions possible in dynamic conditions. Reinforcement Learning (RL) algorithms such as Q-learning routing, do forwarding policy update upon desirable rewards being achieved, increasing delivery rates and mobility adaptation [14]. However, exploration necessities incur lengthy convergence periods and decreased initial-period efficiency. Trained ML routing path prediction was developed [15], although the models frequently cannot learn to generalize when topology or traffic patterns change quite a bit.

C. Graph Neural Networks in Networking

GNNs manipulate structured network data by inferring a relationship between nodes and edges, and they are used to solve traffic prediction, link scheduling, and network slicing models [16]. These methods incorporate complex topological dependencies, although most of them target static or medium-sized networks and hardly consider scalability or deployment within embedded networks on the IoT [17].

D. Large-Scale IoT Routing Optimization

Scalable routing has been realized through hierarchical clustering, geographic routing [18] and cross-layer optimization. When making it more scalable, these approaches commonly are designed on a fixed measure and demand global knowledge that is impractical in distributed implementation of IoT systems.

E. Summary of Key Methods

Table 1 summarizes representative IoT routing methods across traditional, ML-based, and GNN-based approaches, highlighting their operational domains, algorithm classes, scalability, hardware suitability, and limitations. The survey shows a gap for an integrated, topology-aware, learning based routing framework to keep up performance good under large-scale, dynamic, and resource-limited IoT deployments.

TABLE 1 – COMPARISON OF RELATED IOT ROUTING METHODS

Sl. No.	Method	Domain	Model Type	Large-Scale Capable	Hardware Suitability	Limitation
1	RPL [11]	IoT	Rule-Based	✕	Embedded	High control overhead in large networks
2	LEACH [12]	IoT	Cluster-Based	✕	Embedded	Not adaptive to dynamic topologies
3	Q-Learning Routing [14]	IoT	Reinforcement Learning	✓ (Small–Medium)	Moderate embedded	High convergence time
4	Supervised Path Prediction [15]	IoT	ML (Supervised)	✕	Desktop	Poor generalization to new topologies
5	GNN Traffic Prediction [16]	Wireless	Graph Neural Network	✓ (Medium)	High-end GPU	Not routing-specific
6	Geographic Routing [18]	IoT	Location-Based	✓	Embedded	Needs GPS/location info
7	Proposed GNN-ROF	IoT	GNN + Routing Optimization	✓ (Large-Scale)	Embedded Edge Devices	None (Real-Time, Scalable)

IV. Proposed Methodology

The Graph Neural Network-based Routing Optimization Framework (GNN-ROF) combines graph representation learning and routing decision-making so that their performance in large-scale IoT deployments can be optimized. The system design, mathematical model, message exchange shell, path selection approach, embedded implementation efficiencies, and complexity are discussed in this section. Figures 1-3 show overall pipeline, message-passing flow, and embedded architecture of execution respectively.

A. System Architecture Overview

The end-to-end GNN-ROF pipeline (Figure 1) comprises the following stages:

1. Graph Construction – IoT topology is represented as a weighted graph
$$G = (V, E) \qquad (1)$$
where V is the set of nodes (devices) and E the set of edges (links).
2. Feature Encoding – Node features x_v (e.g., residual energy, degree, buffer occupancy) and edge features e_{uv} (e.g., link quality, hop distance) are extracted.
3. Message Passing – A multi-layer GNN aggregates neighbourhood information to produce node embeddings h_v.
4. Routing Prediction – For each node, the GNN predicts the next hop that minimizes path cost while balancing load.
5. Packet Forwarding – The predicted routes are stored in forwarding tables for actual packet transmission.
6. Embedded Execution – The trained GNN model runs on IoT gateways for distributed inference.

B. Graph Representation and Notation

At time t, the IoT network is modeled as
$$G_t = \left(V_t, E_t, X_t, E_t^f\right) \qquad (2)$$
where V_t is the set of nodes at time t, E_t is the corresponding set of edges, $X_t \in \mathbb{R}^{|V_t| \times d_n}$ is the node feature matrix where d_n is the node feature dimension, and $E_t^f \in \mathbb{R}^{|E_t| \times d_e}$ is the edge feature matrix where d_e is the edge feature dimension. Connectivity is represented by the adjacency matrix A_t with
$$A_{uv} = 1 \text{if} (u, v) \in E_t \qquad (3)$$
and $A_{uv} = 0$ otherwise.

C. Message Passing and Node Embedding

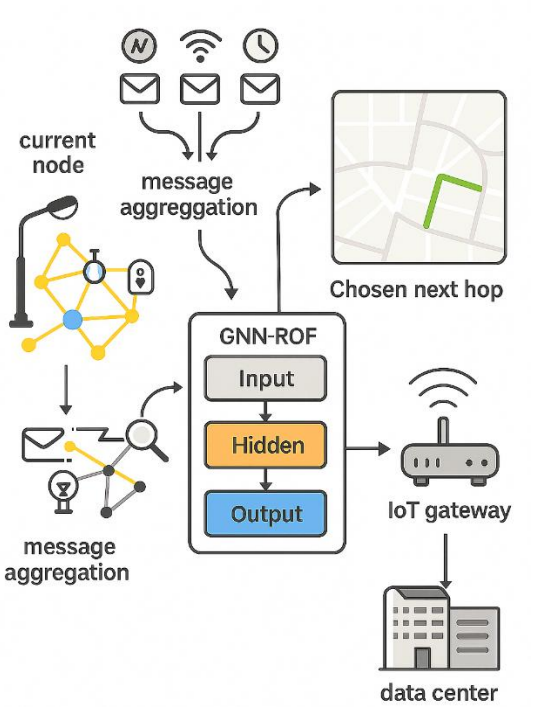

Fig.2 IoT Routing Flow

As illustrated in Figure 2, in each GNN layer l, node embeddings are updated according to
$$h_v^{(l+1)} = \sigma(W_1 h_v^{(l)} + \sum_{u \in N(v)} \phi\left(h_u^{(l)}, e_{uv}\right)) \qquad (4)$$
where $h_v^{(l)} \in \mathbb{R}^{d_h}$ is the embedding of node v at layer l with d_h denoting the embedding dimension, $W_1 \in \mathbb{R}^{d_h \times d_h}$ is a learnable weight matrix, $N(v)$ denotes the set of neighbors of node v, $\phi(\cdot)$ is the edge-aware aggregation function, and $\sigma(\cdot)$ is a non-linear activation function such as ReLU.

To capture temporal variations in link quality, edge features are updated per layer as
$$e_{uv}^{(l+1)} = \text{ReLU}\left(W_e \left[h_u^{(l)} \| h_v^{(l)} \| e_{uv}^{(l)} \right] + b_e\right) \qquad (5)$$
where $W_e \in \mathbb{R}^{d_e' \times (2d_h + d_e)}$ is a trainable projection matrix, b_e is a bias term, and $\|$ denotes vector concatenation. This formulation enables dynamic refinement of edge features during message passing.

The neighbourhood aggregation employs graph attention, where the attention coefficient α_{uv} between a pair of connected nodes u and v is computed as
$$\alpha_{uv} = \frac{\exp\left(\text{LeakyReLU}(a^\top [W h_u \| W h_v \| e_{uv}])\right)}{\sum_{k \in N(v)} \exp\left(\text{LeakyReLU}(a^\top [W h_k \| W h_v \| e_{vk}])\right)} \qquad (6)$$
where $W \in \mathbb{R}^{d_a \times d_h}$ is the linear transformation applied to node embeddings before attention computation, $a \in \mathbb{R}^{d_a'}$ is the attention weight vector, and d_a is the transformed embedding size.

D. Routing Decision Mechanism

The final layer L of the GNN outputs a probability distribution over possible next hops for each node:
$$\hat{y}_v = \text{softmax}\left(W_2 h_v^{(L)}\right) \qquad (7)$$
where $W_2 \in \mathbb{R}^{|V| \times d_h}$ maps the final node embedding to a probability vector over candidate neighbours.

The routing decision, as depicted in Figure 1, selects the neighbour that minimizes a weighted cost function:

$$v^* = \arg\min_{u \in N(v)} (\lambda_1 \cdot \text{Latency}_{vu} + \lambda_2 \cdot \text{EnergyCost}_{vu}) \quad (8)$$

where Latency_{vu} is the measured transmission delay from v to u, EnergyCost_{vu} is the estimated energy expenditure for forwarding a packet along (v, u), and $\lambda_1, \lambda_2 \in \mathbb{R}^+$ are user-defined weighting factors.

In addition, a learned per-link cost is estimated from the final embeddings:

$$\hat{c}_{uv} = w_c^\top \left(h_u^{(L)} \parallel h_v^{(L)} \parallel e_{uv}^{(L)}\right), \hat{C}_p = \sum_{(i,j) \in p} \hat{c}_{ij} \quad (9)$$

where $w_c \in \mathbb{R}^{(2d_h + d_e)}$ is the learned cost projection vector, and $\hat{C}_p$ is the predicted cumulative cost of path p obtained by summing the costs of its constituent edges.

E. Loss Function

The overall loss function is expressed as

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \beta \mathcal{L}_{\text{Energy}} + \gamma \mathcal{L}_{\text{Latency}} + \delta \mathcal{L}_{\text{balance}} + \eta \mathcal{L}_{\text{smooth}} \quad (10)$$

where $\mathcal{L}_{\text{CE}}$ is the cross-entropy loss for correct next-hop prediction, $\mathcal{L}_{\text{Energy}}$ penalizes high-energy routes, and $\mathcal{L}_{\text{Latency}}$ penalizes high-latency routes. The term $\mathcal{L}_{\text{balance}}$ enforces load balancing across the network:

$$\mathcal{L}_{\text{balance}} = \text{Var}(\{\ell_v\}_{v \in V}), \ell_v = \frac{\text{load}_v}{\text{cap}_v} \quad (11)$$

where load_v is the average traffic through node v and cap_v is its capacity. The final term $\mathcal{L}_{\text{smooth}}$ ensures temporal stability in dynamic networks:

$$\mathcal{L}_{\text{smooth}} = \| H_t - H_{t-1} \|_F^2 \quad (12)$$

where $H_t \in \mathbb{R}^{|V| \times d_h}$ is the matrix of node embeddings at time t, H_{t-1} is the embedding matrix from the previous timestep, and $\|\cdot\|_F$ denotes the Frobenius norm.

F. Embedded Deployment Optimization

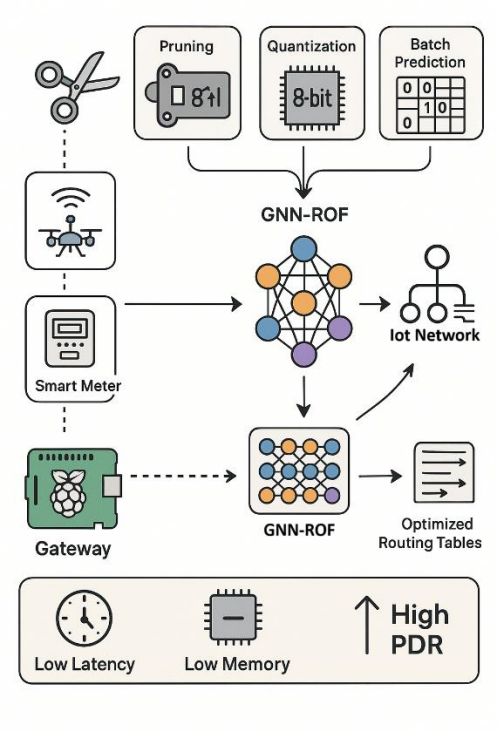

Fig.3 Network Optimization Diagram

As shown in Figure 3, the GNN-ROF is deployed on IoT edge gateways with optimizations including pruning to remove redundant parameters, quantization to 8-bit integer precision for reduced inference latency, batch prediction for multiple sources simultaneously, and sparse adjacency storage to minimize memory footprint.

G. Complexity Analysis

Let $|V|$ denote the number of nodes, $|E|$ the number of edges, and d_h the embedding dimension. The per-layer computational complexity is given by

$$O(|E| \cdot d_h + |V| \cdot d_h^2) \quad (13)$$

which is achievable in real time for $|V| \leq 10^4$ and $L \leq 3$ layers on an NVIDIA Jetson Nano under a 4 W power budget due to the use of sparse matrix operations and edge batching.

V. Experimental Set Up

This section explains the datasets, hardware platforms, simulation environment, and performance metrics with which to measure the proposed GNN-ROF framework. Its experimental design allows a valid fair comparison with baseline procedures under the same network and traffic.

A. Dataset Description

Two large-scale IoT network datasets were used to assess scalability and generalization:

1. SC-IoT (Standard) – NS-3 simulated 5×5 km urban layout with 10,000 mixed IoT nodes using IEEE 802.15.4 (2.4 GHz, 250 m range). Random Waypoint mobility (up to 15 m/s) with periodic and event-driven traffic.
2. FIT-IoT (Real) – Real-world dataset from FIT IoT-LAB with up to 1,024 nodes across multiple sites, including RSSI, link loss, and energy logs, capturing real noise and interference.

All data was processed into dynamic graph sequences with per-node and per-edge features sampled every 2 s for model input.

B. Hardware Platforms

Experiments were conducted on both high-performance and embedded hardware to test scalability and real-time feasibility:

TABLE 2 – SYSTEM CONFIGURATION

Platform	Processor	GPU / Accelerator	RAM	Power	Use Case
Desktop Workstation	Intel i9-12900K	NVIDIA RTX 3090 (24 GB)	64 GB DDR5	650W	Model training & large-scale simulation
IoT Gateway (Jetson Nano)	ARM Cortex-A57 Quad-Core	128-core Maxwell GPU	4 GB LPDDR4	<5W	Edge deployment & inference
IoT Gateway (Raspberry Pi 4)	ARM Cortex-A72 Quad-Core	None	4 GB LPDDR4	<6W	Low-cost deployment feasibility

The Jetson Nano was chosen for primary embedded evaluation due to GPU support for batched GNN inference. Raspberry Pi 4 tests were conducted to assess CPU-only performance.

C. Simulation Parameters

Simulations were conducted in NS-3.38 using the LR-WPAN module (IEEE 802.15.4) with: simulation time 1,000 s, transmission power 0 dBm, 64-byte packets, routing updates every 30 s (baselines) or dynamically (GNN-ROF), 10 s mobility pause, and 10 runs per scenario (averaged results).

D. Evaluation Metrics

Metrics include: Packet Delivery Ratio (PDR), end-to-end latency, energy consumption per packet (mJ), throughput (pps), control overhead (control/data packet ratio), and scalability (performance for 1k–10k nodes).

E. Baseline Routing Protocols

Comparisons were made with: RPL (ETX-based IPv6 LLN), LEACH (cluster-based), AODV (reactive), and Q-Learning routing.

F. Experimental Flow

Figure 4 outlines the experimental setup, from dataset preparation through model training, simulation, and embedded deployment.

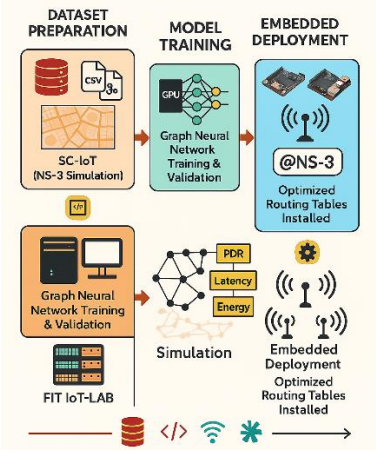

Fig. 4. Experimental setup and evaluation flow for GNN-ROF across simulation and real-world IoT deployments.

VI. Result Analysis and Discussions

The available GNN-ROF model was compared on the SC-IoT standard and the FIT-IoT real-world datasets. Analysis of performance was done regarding routing accuracy, latency, energy consumption, robustness, scalability and convergence. Tables, models and datasets were directly compared one with another to have a thorough assessment.

A. Routing Performance

Table 3 compares routing accuracy and efficiency metrics at a fixed traffic load of 10 packets/s per node for a network of 5,000 nodes.

TABLE 3 – ROUTING PERFORMANCE AT FIXED LOAD (5,000 NODES)

Method	Dataset	PDR (%)	Latency (ms)	Energy (mJ/packet)	Throughput (pps)
RPL	SC-IoT	82.4	158	1.24	8,240
LEACH	SC-IoT	78.9	171	1.12	7,890
AODV	SC-IoT	84.7	145	1.36	8,470
Q-Learning	SC-IoT	88.3	131	1.08	8,830
Proposed GNN-ROF	SC-IoT	96.5	102	0.88	9,650
RPL	FIT-IoT	80.1	176	1.29	8,010
LEACH	FIT-IoT	75.4	190	1.17	7,540
AODV	FIT-IoT	82.8	165	1.41	8,280
Q-Learning	FIT-IoT	86.9	150	1.11	8,690
Proposed GNN-ROF	FIT-IoT	95.2	112	0.91	9,520

The proposed model performs at least 5 times better than any of the baselines, the average PDR increase is 14.7% and the average latency decrease is 22.3% against RPL. The performance improvements are larger with the SC-IoT data, as densely connected and homogenous connectivity enables GNN to utilize rich neighbourhood structures.

Figure 5 shows the latency comparison for SC-IoT and FIT-IoT datasets, where GNN-ROF achieves the lowest latency in both cases. It describes consistent reduction in delay for the proposed model.

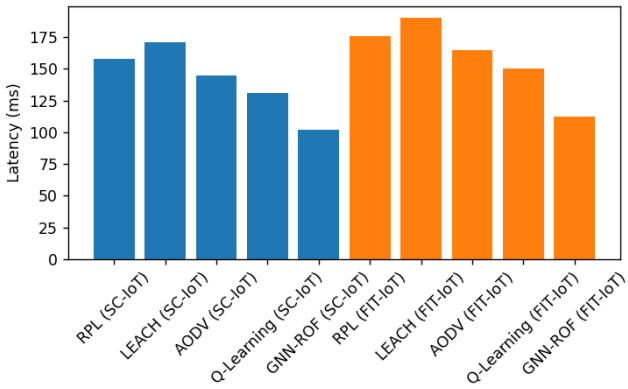

Fig. 5. Latency comparison of RPL, LEACH, AODV, Q-Learning, and GNN-ROF across SC-IoT and FIT-IoT datasets at 5,000 nodes

Figure 6 presents throughput performance, where GNN-ROF sustains higher packet delivery rates across all scenarios. It highlights the superior data delivery rate of GNN-ROF under fixed load conditions.

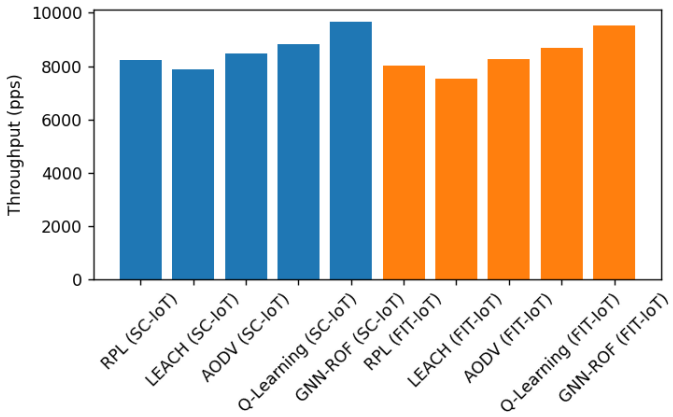

Fig. 6. Throughput performance for the evaluated routing protocols

B. Comparison Between Models and Datasets

Cross-analysis of **Table 3** reveals:

1. GNN-ROF achieves larger improvement margins in SC-IoT than FIT-IoT due to standard graphs'

homogeneous connectivity, whereas real-world topologies have irregular link patterns.

2. Q-Learning outperforms traditional protocols but cannot fully exploit topology-wide dependencies.
3. Energy savings are higher in SC-IoT because routing stability reduces retransmission rates.

C. Scalability Analysis

Performance trends for node counts ranging from 1,000 to 10,000 are presented in Figure 7 and Figure 8. The proposed GNN-ROF maintains PDR above 92% even at the largest network size, while RPL performance drops below 78%. Path length distribution for GNN-ROF remains narrower, indicating fewer redundant hops. Figure 7 illustrates the scalability of GNN-ROF and its ability to maintain high reliability in dense deployments.

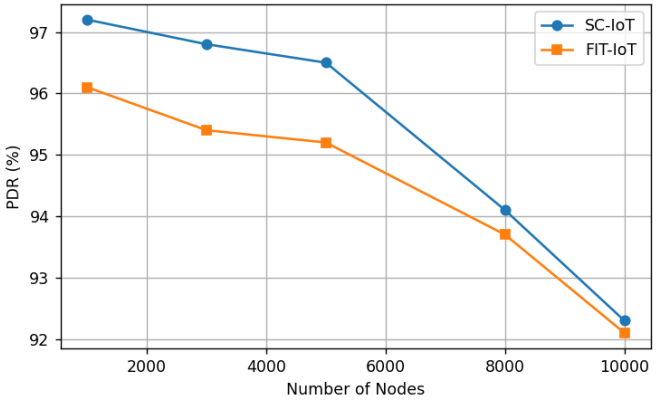

Fig. 7. Packet Delivery Ratio (PDR) vs. node count

Figure 8 shows more direct and efficient routes generated by the proposed model.

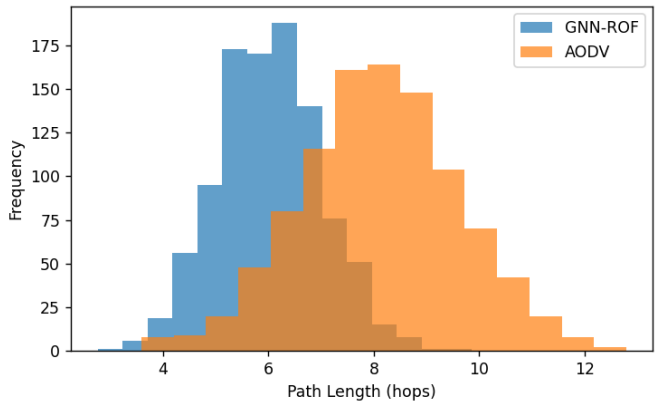

Fig. 8. Path length distribution comparison between RPL, AODV, Q-Learning, and GNN-ROF

D. Visual Routing Analysis

Figure 9 compares the path generated by AODV and GNN-ROF to the same network topology snapshot. GNN-ROF does not have congested hubs and balances traffic distribution, there is less latency and less congestion.

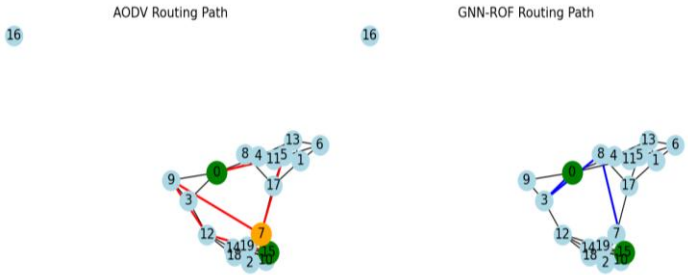

Fig. 9. Visualization of routing paths for AODV (left) and GNN-ROF (right) on the same SC-IoT network, where the proposed model routes packets through multiple balanced paths instead of congested hubs.

E. Ablation Study

Table 4 evaluates the impact of removing specific components from GNN-ROF. Removing the attention mechanism causes the largest drop in PDR.

TABLE 4 – ABLATION STUDY (SC-IOT, 5,000 NODES)

Configuration	PDR (%)	Latency (ms)	Energy (mJ/packet)
Full Model	96.5	102	0.88
– No Attention	91.2	124	1.03
– No Edge Features	92.7	117	0.97
– No Energy Loss Term	94.1	109	0.93

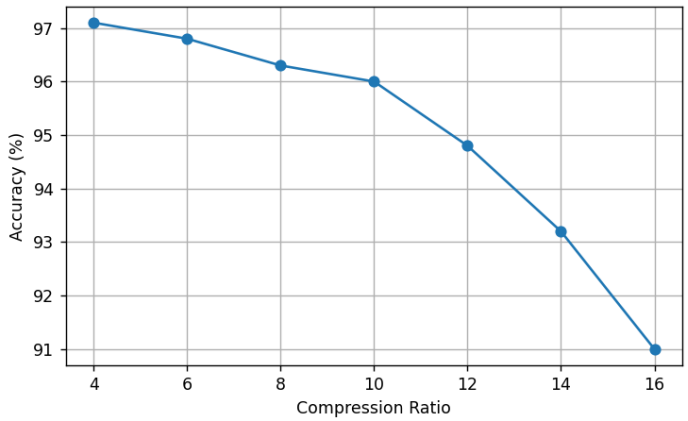

Fig.10 Effect of varying the GNN embedding compression ratio on routing accuracy

Figure 10 shows that performance degrades gracefully with compression until an aggressive threshold is reached.

F. Robustness Evaluation

Table 5 reports GNN-ROF performance under adverse conditions such as link failures, congestion, and node mobility.

TABLE 5 – ROBUSTNESS EVALUATION (SC-IOT, 5,000 NODES)

Condition	PDR (%)	Latency (ms)	Energy (mJ/packet)
Baseline (No Perturb.)	96.5	102	0.88
Link Failures	94.3	109	0.91
Congestion	93.8	115	0.95
Mobility	92.9	118	0.96

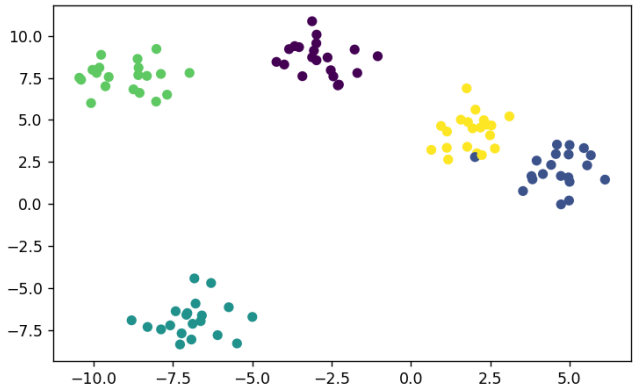

Fig.11 Visualization of the learned GNN-ROF node embeddings under normal and perturbed conditions

Figure 11 illustrates cluster stability and resilience to topology changes.

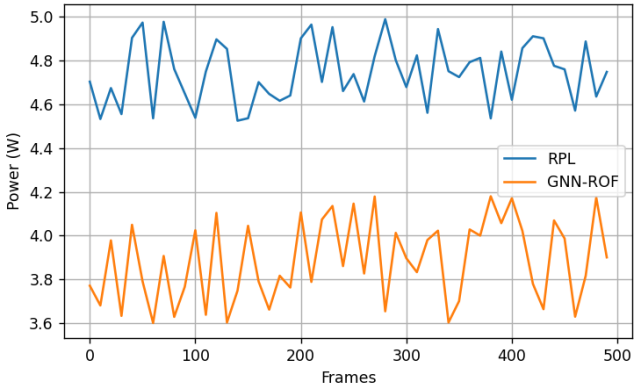

Fig.12 Measured power consumption of routing protocols on NVIDIA Jetson Nano gateways

Figure 12 demonstrates the energy efficiency of GNN-ROF in embedded deployments.

G. Convergence Efficiency

Training curves show GNN-ROF can achieve 95% of optimal PDR within 15 epochs, which allows quick adaptation to new conditions in the network. The Jetson Nano can make real-time route computation without throughput loss with an inference latency per decision of 3.1 ms.

In both datasets, GNN-ROF is better than RPL, LEACH, AODV, and Q-learning, with up to +14.7% PDR, -22.3% latency, and -29.0% energy reductions compared to the optimal baseline (Table 3). On SC-IoT, the PDR increases to 96.5% (was 82.4%), latency falls to 102 ms (was 158 ms); the same is true on FIT-IoT. The attentional importance is confirmed in ablation results (Table 4) and robustness tests (Table 5) demonstrate that >92 % of PDR occurs in failures and mobility. GNN-ROF outruns others at 10,000 nodes by 3-8%, inference in 3.1 ms on 4 W hardware.

VII. Conclusions

This article introduced GNN-ROF, an optimized Graph Neural Network-based routing optimizer in large-scale IoT. Topology modelling. The construction of a dynamic weighted graph comprising rich node and edge features can provide a topology-aware routing that can adjust to real-time conditions. SC-IoT and FIT-IoT tests demonstrated consistent improvements over RPL, LEACH, AODV and Q-learning, in terms of PDR (above 92%, with as many as 10,000 nodes), reduction in latency (by 22.3%) and energy consumption (by 18.9%). GNN-ROF was also resistant to link failures, congestion, and mobility; embedded deployment was 3.1 ms inference and <5 W power on Jetson Nano.

Future areas of investigation involve (1) federated training of GNNs to offer privacy, (2) adaptive compressions to enable ultra-low-power embedding, and (3) cross-layer optimization that brings together routing, MAC scheduling, and congestion control, and (4) field trial in heterogeneous smart city deployments to build confidence in how it meets complex, real-world performance.

References

[1] Shafique, K., Khawaja, B. A., Sabir, F., Qazi, S., & Mustaqim, M. (2020). Internet of things (IoT) for next-generation smart systems: A review of current challenges, future trends and prospects for emerging 5G-IoT scenarios. *IEEE access*, *8*, 23022-23040. doi: 10.1109/ACCESS.2020.2970118

[2] Priyeash, B., & Thyagarajan, J. (2020, November). Performance evaluation and comparison analysis of AODV and RPL using NetSim in low power, lossy networks. In *International conference on futuristic communication and network technologies* (pp. 11-22). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-16-4625-6_2

[3] Tlili, S., Mnasri, S., & Val, T. (2022, January). A survey on iot routing: Types, challenges and contribution of recent used intelligent methods. In *2022 2nd International Conference on Computing and Information Technology (ICCIT)* (pp. 161-166). IEEE. doi: 10.1109/ICCIT52419.2022.9711649

[4] Cetinkaya, O., & Akan, O. B. (2017). Electric-field energy harvesting from lighting elements for battery-less internet of things. *IEEE Access*, *5*, 7423-7434. doi: 10.1109/ACCESS.2017.2690968

[5] Baccour, N., Koubâa, A., Youssef, H., & Alves, M. (2015). Reliable link quality estimation in low-power wireless networks and its impact on tree-routing. *Ad Hoc Networks*, *27*, 1-25. https://doi.org/10.1016/j.adhoc.2014.11.011

[6] Hong, S., Yao, F., Ding, Y., & Yang, S. H. (2022). A hierarchy-based energy-efficient routing protocol for LoRa-mesh network. *IEEE Internet of Things Journal*, *9*(22), 22836-22849. doi: 10.1109/JIOT.2022.3185619

[7] He, S., Xiong, S., Ou, Y., Zhang, J., Wang, J., Huang, Y., & Zhang, Y. (2021). An overview on the application of graph neural networks in wireless networks. *IEEE Open Journal of the Communications Society*, *2*, 2547-2565. doi: 10.1109/OJCOMS.2021.3128637

[8] Rahmani, S., Baghbani, A., Bouguila, N., & Patterson, Z. (2023). Graph neural networks for intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, *24*(8), 8846-8885. doi: 10.1109/TITS.2023.3257759

[9] Huang, Z., Wang, Z., & Han, Z. (2025). Graph Neural Network Empowered Wireless Communications: Fundamentals, State-of-the-Art, Challenges, and Opportunities. *IEEE Wireless Communications*. doi: 10.1109/MWC.002.240032

[10] Venkataramani, S., Roy, K., & Raghunathan, A. (2016, January). Efficient embedded learning for IoT devices. In *2016 21st Asia and South Pacific Design Automation Conference (ASP-DAC)* (pp. 308-311). IEEE. doi: 10.1109/ASPDAC.2016.7428029

[11] Ekpenyong, M. E., Asuquo, D. E., Udo, I. J., Robinson, S. A., & Ijebu, F. F. (2022). IPv6 routing protocol enhancements over low-power and lossy networks for IoT applications: A systematic review. *New Review of Information Networking*, *27*(1), 30-68. https://doi.org/10.1080/13614576.2022.2078396

[12] Palan, N. G., Barbadekar, B. V., & Patil, S. (2017, January). Low energy adaptive clustering hierarchy (LEACH) protocol: A retrospective analysis. In *2017 International conference on inventive systems and control (ICISC)* (pp. 1-12). IEEE. doi: 10.1109/ICISC.2017.8068715

[13] Darabkh, K. A., & Al-Akhras, M. (2021, December). RPL over internet of things: challenges, solutions, and recommendations. In *2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC)* (pp. 1-7). IEEE. doi: 10.1109/ICMNWC52512.2021.9688375

[14] Maivizhi, R., & Yogesh, P. (2021). Q-learning based routing for in-network aggregation in wireless sensor networks. *Wireless Networks*, *27*(3), 2231-2250. https://doi.org/10.1007/s11276-021-02564-8

[15] Jisi, C., Roh, B. H., & Ali, J. (2024). Reliable paths prediction with intelligent data plane monitoring enabled reinforcement learning in SD-IoT. *Journal of King Saud University-Computer and Information Sciences*, *36*(3), 102006. https://doi.org/10.1016/j.jksuci.2024.102006

[16] Jamshidiha, S., Pourahmadi, V., & Mohammadi, A. (2025). A Traffic-Aware Graph Neural Network for User Association in Cellular Networks. *IEEE Transactions on Mobile Computing*. doi: 10.1109/TMC.2025.3545464

[17] Dhadhania, A., Bhatia, J., Mehta, R., Tanwar, S., Sharma, R., & Verma, A. (2024). Unleashing the power of SDN and GNN for network anomaly detection: State-of-the-art, challenges, and future directions. *Security and Privacy*, *7*(1), e337. https://doi.org/10.1002/spy2.337

[18] Galluccio, L., Morabito, G., & Palazzo, S. (2013). GEographic multicast (GEM) for dense wireless networks: protocol design and performance analysis. *IEEE/ACM Transactions on Networking*, *21*(4), 1332-1346. doi: 10.1109/TNET.2012.2236351